What is claimed is:

1. A method of identifying one or more marker genes whose level of expression predicts a biological response in a cell or tissue, comprising:

(a) obtaining gene expression data for at least a test cell or tissue group and a control cell or tissue group, each of the test and control groups comprising a collection of cell or tissue samples;

(b) analyzing the gene expression data to identify said one or more marker genes by a linear discriminate metric that does not include a variable corresponding to a factor selected from the group consisting of: the magnitude of the difference in gene expression for each gene between the test and control groups, and the behavior or identity of other genes detected in the samples or across samples and combinations thereof, thereby identifying one or more marker genes whose level of expression predicts a biological response in a cell or tissue.

2. A method of claim 1, wherein the test cell or tissue group is exposed to a toxin.

3. A method of claim 1, wherein step (b) does not rely on a correlation matrix to account for relationships or interdependencies between the genes.

4. A method of claim 2, wherein step (b) comprises computing probability distributions from the gene expression data for each gene found in the toxin exposed cell or tissue group and the control cell or tissue group.

5. A method of claim 2, wherein the toxic response is hepatotoxicity.

6. A method of claim 2, wherein the toxin and control groups comprise at least about 10 samples.

7. A method of claim 2, wherein the toxin and control groups comprise at least about 50 samples.

8. A method of claim 2, wherein each gene is given equal weight as a predictive marker.

9. A method of claim 1, wherein step (b) comprises determining a scoring function and a discriminate score for each gene.

10. A method of claim 9, wherein the scoring function $f_1(Z)$ is generated according to the formula:

$$f_1(z) = \frac{\left(1/\sqrt{v1}\right)e^{-1/2(z_i-m1)^2/v1}}{\left(\left(1/\sqrt{v1}\right)e^{-1/2(z_i-m1)^2/v1} + \left(1/\sqrt{v2}\right)e^{-1/2(z_i-m2)^2/v2}\right)}$$

wherein Z is a cell or tissue sample;

$f_1(Z)$ is the discriminate score for sample Z relative to the samples of group 1;

m1 is an estimator of the true mean for the distribution of gene expression levels for the samples of group 1;

m2 is an estimator of the true mean for the distribution of gene expression levels for the samples of group 2;

v1 is an estimator of the variance in the distribution of gene expression levels for the samples of group 1; and

v2 is an estimator of the variance in the distribution of gene expression levels for the samples of group 2.

11. A method of claim 10, wherein a discriminate score for each gene is calculated by summing the instances where the scoring function returns a value for each sample that is greater than a predetermined significance level.

12. A method of claim 11, wherein the discriminate score (P) for each gene is calculated between the test and control groups by:

(a) counting the number of times, P1, that $f1(Xi) > ½$ for i =1..t and the number of times, P2, that $f1(Yj) < ½$ for j = 1..n; and

(b) calculating the P discriminate score according to the formula:

$(P1 + P2)/(n + t)$.

13. A method of claim 1, wherein the one or more marker genes discriminate between the test and control groups.

14. A method of claim 1, wherein the gene expression data comprises measurement of the amount or relative amount of mRNA for each gene.

15. A method of claim 1, wherein the one or more markers are cross-validated.

16. A method of claim 1, wherein the markers stratify a patient population.

17. A method of claim 16, wherein the patient population is stratified for its response to the administration of a drug or according to a physiological characteristic relevant to a clinical trial.

18. A method of building a classification database, comprising:

(a) obtaining gene expression data for at least one test group of cell or tissue samples and at least one control group of cell or tissue samples;

(b) identifying marker genes by an LDA metric capable of discriminating between the test and control groups and storing these markers in the database;

(c) evaluating the predictive ability of at least one of the markers of step (b); and

(d) refining the database with information derived from additional biological studies.

19. A method of claim 18, wherein the database comprises gene expression information and meta-data.